Methods For Protein Analysis

Edited By
John P. Cherry
Robert A. Barford

Eastern Regional Research Center
Agricultural Research Service
U.S. Department of Agriculture
600 E. Mermaid Lane
Philadelphia, PA 19118

American Oil Chemists' Society
Champaign, Illinois

1988

Chapter Fourteen

Fourier Transform Infrared Spectroscopy in Protein Conformation Studies

Heino Susi* and D. Michael Byler

Eastern Regional Research Center
ARS, U.S. Department of Agriculture
600 E. Mermaid Lane
Philadelphia, PA 19118

Abstract

Fourier transform infrared spectroscopy permits more detailed studies of the secondary structure of proteins than older, dispersive methods. The principle reason is the very high signal-to-noise ratio which can be obtained, frequently better than 1000:1. Best results are obtained by applying mathematical resolution/enhancement techniques such as Fourier self-deconvolution and second derivative spectroscopy. Both qualitative and quantitative studies can be carried out. The results are more specific than the ones obtained by other spectroscopic techniques.

Introduction

Until recently, infrared spectroscopy has been essentially a qualitative tool for conformational studies of proteins, although some semi-quantitative applications have been attempted with the help of digital computers (1-3). The use of Fourier transform infrared spectroscopy (FTIR) has led to many improvements in both qualitative and quantitative applications (4,5). The two main advantages of Fourier transform infrared spectroscopy over the older dispersive methods are first, the higher signal-to-noise (S/N) ratio (which facilitates the observation of very small absorbance changes) and second, better wavelength accuracy and reproducibility. For the purpose of protein structure studies, the first advantage (high S/N ratio) is of greatest importance because it permits the use of completely new techniques for extracting a maximum of structural information from the spectra. The most important of these new techniques are resolution enhancement through Fourier self-deconvolution (5,6), resolution enhancement through second deriva-

*Deceased.

tive spectroscopy (7,8), and quantitative secondary structure (conformation) analysis by combining self-deconvolution with curve fitting (9,12). The application of deconvolution and curve fitting to the spectra of the proteins ribonuclease A and α-chymotrypsin is illustrated in Figure 14.1. The upper curves (crosses) give the digitized FTIR spectra before deconvolution. In the lower curves, the crosses represent the experimental data after deconvolution. The individual peaks in the lower curves (solid lines) are the components determined by curve fitting while the solid lines, which closely fit the deconvolved experimental spectra, show the summation of these components. It is evident that by using these techniques, ill-defined broad absorption bands can be resolved into components with well-defined frequencies and areas (integrated intensities).

How second derivative spectroscopy is applied to the spectrum of the

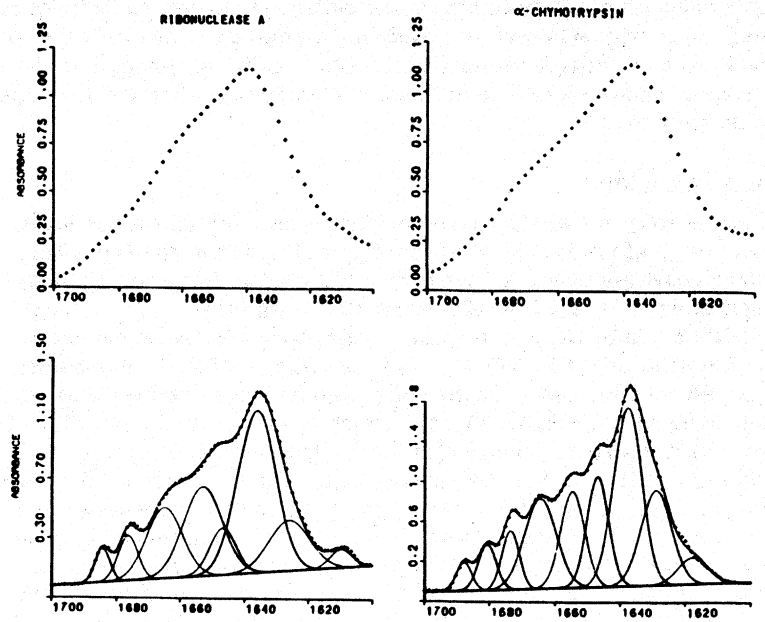

Fig. 14.1. Original FTIR spectrum (top) and deconvolved FTIR spectrum (bottom) of ribonuclease A and α-chymotrypsin in the 1600-1700 cm^{-1} region. Absorbance plotted vs. wavenumber (cm^{-1}). Crosses: digitized original spectra and digitized deconvolved spectra. Solid lines: calculated individual component-bands and calculated overall deconvolved spectrum. 5% (w/v) protein in D_2O with 0.01M NaCl; pD 7; pathlength = 0.075 mm.

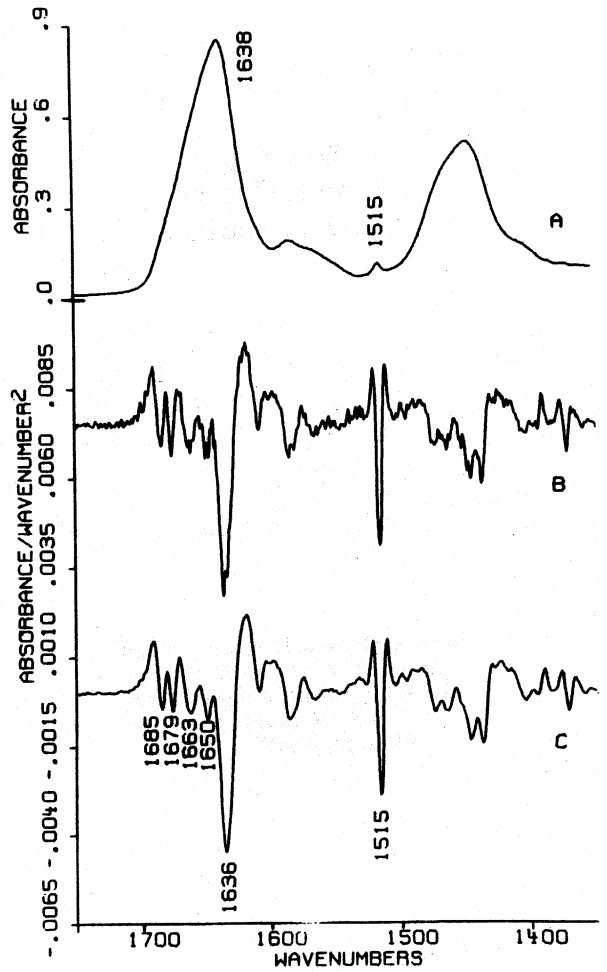

Fig. 14.2. (A) Original FTIR spectrum of ribonuclease A. (B) Unsmoothed second derivative spectrum of ribonuclease A. (C) Second derivative spectrum of ribonuclease A subjected to 9-point smoothing. 5% w/v in D_2O; pathlength = 0.075 mm; pD 7. [From ref. 8]

protein ribonuclease A is shown in Figure 14.2. The top curve gives the original spectrum, the middle curve shows the "unsmoothed" second derivative, and the bottom curve shows the "smoothed" second derivative spectrum. The frequencies in the 1600 to 1700 cm^{-1} range of the second derivative spectrum of ribonuclease A agree well with corresponding frequencies obtained by self-deconvolution as shown in Figure 14.1. This agreement demonstrates how well the two techniques complement each other.

As demonstrated by the data in Figures 14.1 and 14.2, the broad infrared bands of proteins, which result from the vibrations of the peptide groups, are actually composed of several components; these are associated with the helical segments, extended chains and "turns" into which the peptide backbone folds. The most diagnostic of the several infrared bands of proteins is the so-called *amide I* band in the region of 1600–1700 cm^{-1}. Because the inherent band widths of the individual components are much greater than the spectral resolution of most instruments (10–20 cm^{-1} as compared to 0.1–2 cm^{-1}, respectively), the observed original bands cannot be optically resolved into distinct components associated with specific classes of protein structure. As shown in Figures 14.1 and 14.2, resolution is nevertheless achievable by means of mathematical techniques such as second derivative spectroscopy and Fourier self-deconvolution.

Because these resolution enhancement methods greatly accentuate the intensity of weak but narrow noise spikes relative to the much broader protein bands, reliable results can be obtained only if the original spectrum has a sufficiently high S/N ratio. This is much more easily achieved by using modern FTIR interferometers than with older dispersive instrumentation.

Theoretical

General. A Fourier transform infrared spectrometer is fundamentally different from the older dispersive instruments. In the latter, a grating or prism disperses a collimated beam of infrared light onto a slit which effectively blocks all but a narrow range of frequencies from reaching the detector. By continuously changing the angle of the grating with respect to the incident light beam, a complete spectrum can be scanned, one spectral resolution element at a time. The FTIR instrument, by contrast, is nondispersive and makes use of an interferometer simultaneously to encode data from the whole spectral range. In general, the interferometer is some variation of the original design by Michelson.

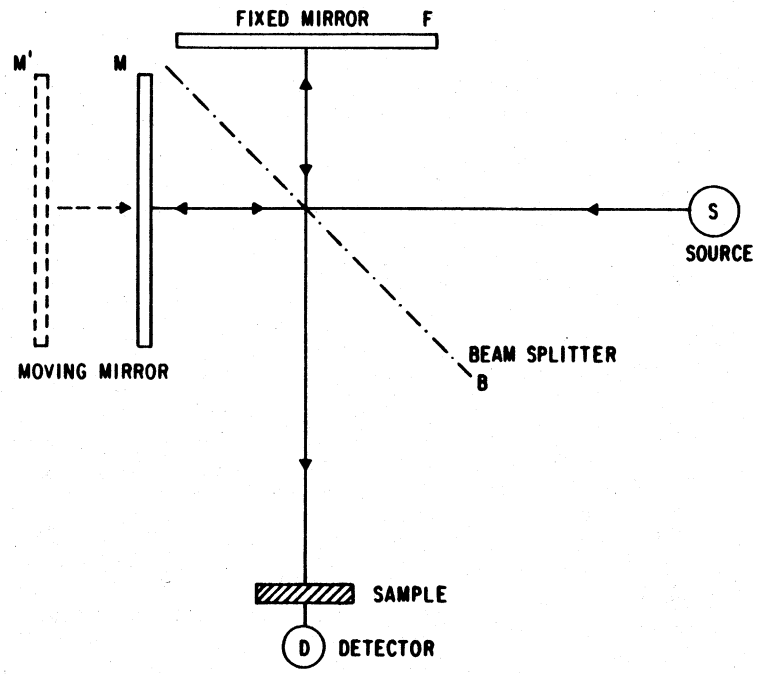

Fig. 14.3. Schematic representation of a Michelson interferometer.

The principal features of a typical Michelson interferometer are depicted in Figure 14.3. Set at right angles to one another is a pair of plan mirrors, F and M; mounted between is a semi-reflecting film, or beamsplitter, B, with its plane at a 45° angle to the mirror faces. As a collimated beam of light from the source S impinges on the beamsplitter B, half is reflected to mirror F and half is transmitted to mirror M. After reflection at the mirrors, the two beams reconverge at the beamsplitter B, where again each is 50% reflected and 50% transmitted. For simplicity, first consider a monochromatic light source of wavelength λ. Twice the difference in the distances from the beamsplitter B to each of the two mirrors is designated as x, the optical retardation, or optical path difference. If x is zero or an integral multiple of λ, the two beams will recombine at B in phase. Due to construc-

tive interference, the signal at the detector will be of maximum intensity. On the other hand, if x has any other value, the two beams of light will be partially out of phase, resulting in destructive interference and decreased detector signal. When $x = (n + 1/2)\lambda$ ($n = 0, 1, 2...$), the signal is zero. If mirror F is fixed in position while mirror M moves at a constant velocity v through some distance r, the signal observed at the detector will be a cosine wave whose frequency is represented by $f = v/\lambda = vk$. (Here k is the wavenumber of the incident radiation and is given in units of cm^{-1}). The amplitude or intensity of this signal as a function of x is I(x) and is called an interferogram. For monochromatic light, I(x) is a pure cosine wave proportional to $\cos(2\mu k \cdot x)$.

In the case of polychromatic light, all frequencies will be in phase only at zero path difference ($x = 0$). At all other values of x, varying degrees of destructive interference will occur. Now the interferogram resulting from one sweep of the moving mirror M will be proportional to the sum of the cosine waves, $\Sigma A_i \cos(2\mu k_i x)$. A_i is the maximum amplitude of the cosine for each incident wavenumber, k_i. (For a frequency continuum, i becomes infinite, and the sum is replaced by an integral.) This interferogram I(x) now has maximum amplitude at $x = 0$ known as the "center burst." If certain frequencies of the incident radiation are absorbed by a sample, the interferogram changes because the amplitudes of the cosine waves of the absorbed frequencies decrease. Even when such changes are readily apparent upon visual inspection of the interferogram, they are difficult to interpret. The interferometric data from the "time domain" are, therefore, Fourier transformed into the "frequency domain" to give an uncorrected, "single-beam" absorption spectrum of the sample, E(k). Dividing E(k) by the spectrum of the incident beam with no sample present, $(E(k)_o$, one obtains the spectrum of the sample in percent transmittance. This is commonly called a "ratioed" spectrum.

FTIR spectrometry offers several theoretical advantages over dispersive infrared measurements (4): First, because the throughput of the incident light is not slit-limited, FTIR spectrometers are inherently more sensitive than dispersive instruments (*Jacquinot's advantage*). For a given source more energy will reach the detector, resulting in a greater signal-to-noise (S/N) ratio. (This advantage is somewhat offset because dispersive intruments can use slow-response thermocouple detectors which have higher detectivities than the most sensitive, low-temperature FTIR detectors for the mid-infrared region.) Second, FTIR spectrometers simultaneously encode all spectral frequencies to give a complete spectrum in a matter of seconds (*Fellgett's or multiplex advantage*). Because the S/N ratio is proportional to the square root of the number of scans, signal-averaging the data from a large number of scans can significantly increase the S/N ratio within a relatively short time. In addition, rapidly occurring spectral changes may be kinetically monitored rather easily. Furthermore, modern, high-speed FTIR spectrometers contain a laser reference interferometer to facilitate digitization of data and for frequency calibration. Because the laser frequency is known to at least seven significant figures, its interferogram frequency is used to calibrate the frequencies of the digitized infrared data points to better than 0.01 cm^{-1} (*Conne's advantage*). The peak positions of observed infrared bands of condensed-phase samples cannot generally be measured with such accuracy, particularly if the bands overlap or are broad. Use of Fourier self-deconvolution and of second derivatives, however, can substantially increase the accuracy of the measured peak positions. Finally, because such good S/N ratios and high wavenumber precision are possible, FTIR spectra can be easily and rapidly manipulated by a computer. This facilitates such mathematical data treatments as interactive spectral subtraction, calculation of second derivatives, Fourier self-deconvolution, and curve fitting.

Second derivative spectra. The second derivative (A''_n) may be calculated by a straightforward analytical method (7,8):

$$A''_n = [A_{n+1} - 2A_{n-1}]/\Delta W^2 \qquad [1]$$

where A_n = absorbance at data point n, $\Delta W = \Delta k/2^m$, Δk = nominal instrumental resolution (cm^{-1}), m = number of times the interferogram is zero-filled prior to Fourier transformation.

The intrinsic shape of a single infrared absorption line may be approximated by a Lorentzian function (6,7):

$$A = \frac{\sigma}{\mu(\sigma^2 + k'^2)} = \frac{1}{\sigma\mu(1 + Bk'^2)} \qquad [2]$$

where A is the absorbance, 2σ is the width at half height, $k'(=k - k_o)$ is the frequency referred to the band center at k_o and $B = 1/\sigma^2$.

The peak frequency for the second derivative is identical to the frequency of the original band center, k_o. The half-width of the second derivative, σ^{II}, is related to the half-width of the original line by (7) $\sigma^{II} = \sigma/2.7$ and the peak intensity of the second derivative, A''_o, to that of the original intensity A_o by $A''_o = -2A_o/\sigma^2$.

To summarize, the second derivative spectrum gives sharp negative peaks with weaker, positive lobes on either side. The peak height of the second derivative is proportional to the original peak height and inversely proportional to the square of the original half-width. Thus, weak but sharp lines, such as arise from water vapor, noise or interference fringes, are greatly accentuated relative to the much broader lines of a condensed phase sample. For real spectra with overlapping bands which deviate from Lorentzian shape, the relationships are more complex. Nonetheless, the above formulae do provide a reasonable approximation for interpreting second derivative spectra (8,9).

Fourier self-deconvolution. The Fourier transformation of a Lorentzian band (Fig. 14.4a) of half-width at half-height, $\sigma(cm^{-1})$, results in a time-domain function, or interferogram, I(x), which is an exponentially-damped cosine wave (Fig. 14.4a) (6,10). The envelope of this function is a decay curve whose exponent is directly proportional to σ. If I(x) is multiplied by the product of an apodization function D(x) and an exponentially-increasing weighting function, the exponent of the decay curve and, thus, the rate of decay of I(x) will be decreased. Therefore, when the reverse Fourier transform is calculated, the width of the resulting spectral band, $E_L(k)$, will be less than that of the original undeconvoluted band, E(k) (Figs. 14.4b-4d). In Figure 14.4, $\tilde{\nu}$ is used for the wavenumber instead of k.

If the positive exponent of the weighting function has the same absolute value as the negative exponent in the decay function, the resulting interferogram $I_L(x)$ will be a cosine wave truncated at x = L. The value of L for $I_L(x)$ is determined by the point L where the apodization function D(x) goes to zero. The instrument resolution, Δk, at which the original spectrum was measured places an approximate upper limit on L of $1/\Delta k$. The reverse Fourier transform of $I_L(x)$ gives a sinc function (sinc x = [sin x]/x) (Fig. 14.4b). The characteristic side-lobes found on either side of the central maximum of this function clearly show that resolution enhancement of bands beyond $L = 1\Delta k$ is not possible, even for perfectly noise-free data. Indeed, to minimize the intensities of side-lobes in the deconvolved spectrum, the boxcar function,

$$D(x) = \begin{cases} 1, \text{ for } x \leq L \\ 0, \text{ for } x > L \end{cases} \qquad [3]$$

is usually replaced by some alternative apodization function (7). These functions reduce the size of the side-lobes at the cost of decreased resolution enhancement. Yang and Griffiths suggest as a rule-of-thumb that,

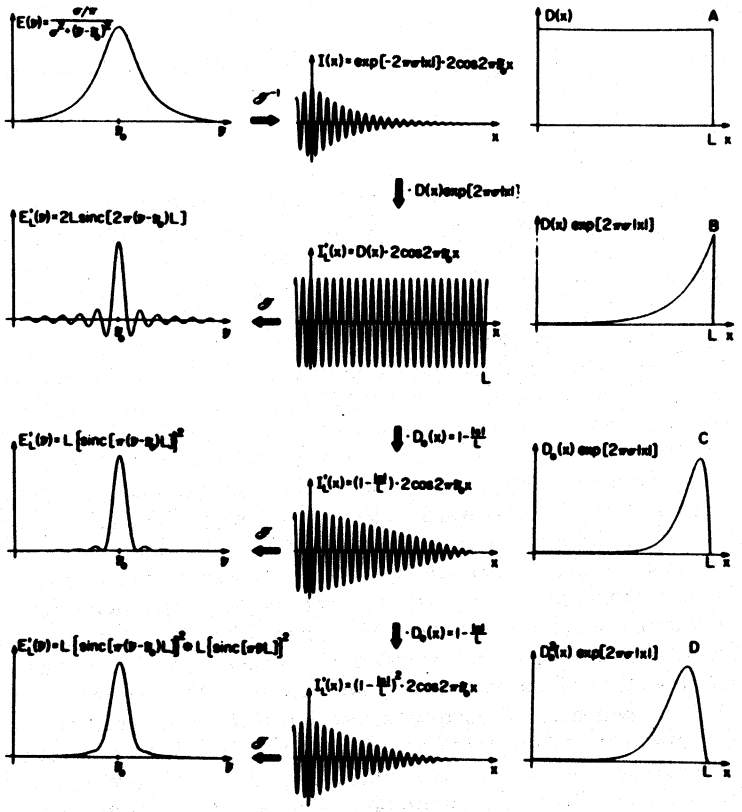

Fig. 14.4 Illustration of the various steps of the Fourier self-deconvolution procedure starting with a Lorentzian line at $\tilde{\nu}_o$ and using different apodization functions, $D_g(x)$: in row (B), $D_g(x)$ equals D(x), the "boxcar function" (see eq. [3] of text); in row (C), $D_g(x)$ equals $D_a(x)$, the "triangular function"; in row (D), $D_g(x) = D^2_a(x)$. Except for row (A), the middle column shows the interferograms $I_L'(x)$, resulting from the application of the functions, $D_g(x)\cdot\exp(2\pi\tilde{\nu}_o x)$ (right-hand column), to the interferogram I(x). These functions are all scaled to the same height. In the left-hand column, except for row (A), the lineshapes E_L ($\tilde{\nu}$), resulting from the self-deconvolution are shown. (6) Note: throughout the text, k is used for wavenumber instead of $\tilde{\nu}$. [From ref. 6, reprinted with permission].

at best, the width of bands after deconvolution cannot be reduced to less than about 1.5(Δk) (10).

Applications

Protein spectroscopy by FTIR in H_2O and D_2O solutions. Infrared spectroscopy in aqueous solution has always been very difficult because water absorbs strongly in the spectral region around 1600-1700 cm^{-1}, where the most important conformation-sensitive protein band (the C=O stretching or *amide I* band) is found. When FTIR first became available, it was thought that the increased sensitivity would render infrared spectroscopy of proteins feasible even in H_2O solution. Instead of the earlier "differential methods" (2), FTIR permitted one to obtain the solution spectrum and the solvent spectrum separately, and then to subtract the latter from the former, to obtain the spectrum of the pure solute. Unfortunately, neither the subtraction nor differential procedures are as straightforward as they first appear. Whenever a solute is present, changes occur in the frequency, width, and height of the solvent band, even for weakly interacting, low-polarity solvent systems. For water and other polar, hydrogen-bonding solvents, such effects become much more pronounced. When strong bands of the solute and the solvent overlap, as in the case of the protein *amide I* band and the bending mode of water, the residual features of the solvent spectrum which remain after subtraction distort the solute bands so that accurate measurement of their frequency and intensity is no longer possible. Because absorption by water is significantly lower in the *amide II* region (1530-1560 cm^{-1}) (1), only minimal difficulty is encountered with solvent subtraction for these protein bands. Side-chain bands, associated with CH_2, CH_3, COO^-, NH_2, and SH groups can also be easily observed in H_2O solution; the latter, though very weak, falls within the relatively clear spectral region around 2560 cm^{-1}.

Koenig and Tabb (11) have investigated the FTIR spectra of several proteins in aqueous (H_2O) solution. The observed *amide I* frequencies appear uncertain, and are actually in error for ribonuclease, evidently because of the difficulties inherent in subtracting the strong H_2O band which absorbs close to 1640 cm^{-1}. We strongly suggest that protein structure studies based on the important *amide I* band which absorbs at 1620-1680 cm^{-1} be carried out in D_2O solution whenever possible. D_2O has no absorption band in this frequency region (2).

Second derivation FTIR spectra. The ideal intrinsic shape of a single infrared absorption line is approximated by a Lorentzian function, as described in the theoretical section. In the second derivative spectrum of such a line, the peak frequency is identical to the original peak frequency, and the half-width is reduced by a factor of 2.7. The peak height is proportional to the original peak height (with the opposite sign), and inversely proportional to the square of the original half-width. There are side-lobes on either side of the principal peak (7). The net result is a very much sharper negative line accompanied by positive side-lobes. Quantitative information, however, is difficult to obtain at present from second derivative spectroscopy because of the complex patterns created by overlapping negative peaks and positive side-lobes.

Second derivative spectra serve three useful purposes. First, the important *amide I* band, which is most useful in studies of secondary structure can generally be resolved into sharp components which represent structural elements such as helical segments, extended strands, and turns. Second, because it displays much more detail than the original spectrum, the second derivative spectrum can be used as an identifying "fingerprint" for any given protein. Third, side-chain groups can be much more easily identified and investigated in second derivative spectra than in the original spectra. Figure 14.5 illustrates all three points. Curve A, the second derivative spectrum of hemoglobin, has only one strong band in the *amide I* region. The band centers at 1652 cm^{-1} and shows that this protein is overwhelmingly helical (8). Curve B, the second derivative spectrum of β-lactoglobulin, has bands at 1623, 1634, and 1679 cm^{-1} which are characteristic of extended chains; and band close to 1650 cm^{-1} which indicates the presence of some helical regions; and bands at 1667 and 1693 cm^{-1} which are assigned to "turns" (8). Curve C represents denatured β-lactoglobulin. All fine-structure in the 1600-1700 cm^{-1} indicates apparently "orderless" chains where most of the peptide groups are probably solvated by water molecules (2,3,8,12).

The value of second derivative spectra as "fingerprints" is also evident from Figure 14.5. The curves A, B and C clearly differ over the entire range depicted, from 1350 to 1750 cm^{-1}, although the original spectra are quite similar.

Furthermore, many side-chain bands which do not even show up in the original spectra are clearly observed in Figure 14.5. For instance, the weak band close to 1610 cm^{-1} is probably associated with arginine residues, and the very sharp band close to 1515 cm^{-1} is caused by tyrosine residues.

Because the older, dispersive-type instruments yield spectra with a relatively low signal-to-noise ratio, they cannot be used for second

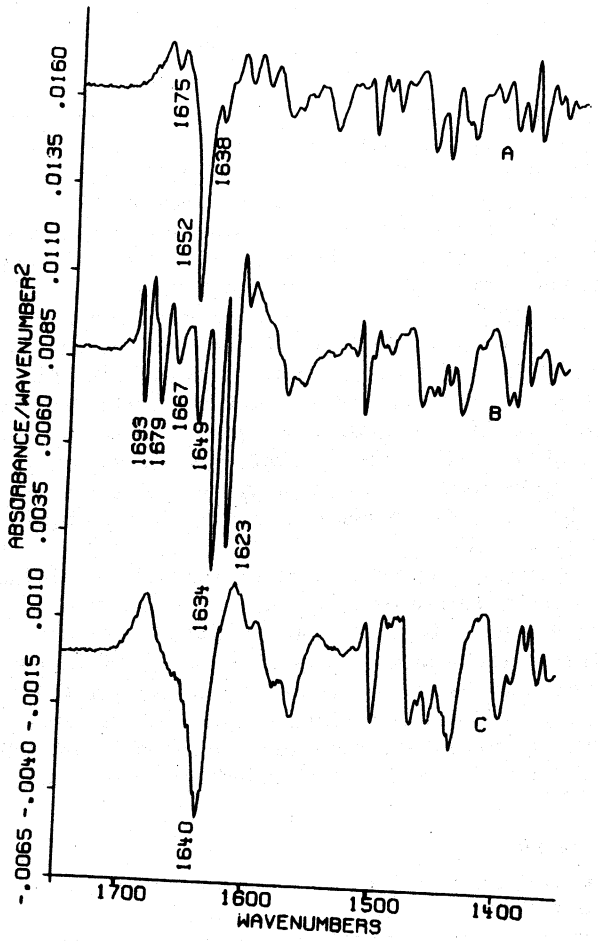

Fig. 14.5. Nine-point smoothed second derivative FTIR spectra: (A) hemoglobin, pD 7; (B) native β-lactoglobulin A, pD 7; (C) denatured β-lactoglobulin, pD 13. 5% w/v in D_2O; pathlength = 0.075 mm. [From ref. 8].

derivative spectroscopy. FTIR, by contrast, can yield spectra with a signal-to-noise ratio of 1000:1 or better and thus permits the observation of second derivative spectra packed with previously unattainable information.

Secondary structure analysis by Fourier self-deconvolution and curve fitting. The self-deconvolved FTIR spectra of ribonuclease A and α-chymotrypsin in the *amide I* region, as obtained in D_2O solution, along with the original spectrum are shown in Figure 14.1. The deconvolved spectra are resolved into Gaussian components by means of a computer program based on Gauss-Newton iteration (9,12). Deconvolution, as carried out by the method described (5,6,9,13), requires two constants as computer input: the estimated half-width of the unresolved component bands, σ, and the resolution "efficiency" factor, K, which reflects the extent of narrowing of the unresolved bands, i.e., the "effective increase in resolution" (6). The choice of these constants is of utmost importance for obtaining meaningful results (5,6,9,12). The constant σ should, ideally, be as close to the true half-width of the unresolved components as possible. When unresolved components with different half-widths must be deconvolved simultaneously, a compromise becomes necessary. This is generally the case for protein spectra. A choice of $\sigma = 6.5$ has generally been found to be satisfactory (5). A maximum value for K can be estimated by the approximate relationship: $k_{max} \approx \log(S/N)$. A K value of 2.4 would thus require a S/N ratio of at least 260. In practice, K values higher than 3 are rarely used. It must be strongly emphasized that an improper selection of σ and/or K can lead to serious errors and artifacts (5). Deconvolution increases with increasing σ, but too large a value of σ leads to "over-deconvolution," resulting in negative side-lobes and distorted spectra. Too high a K value, in turn, leads to excessive noise and spurious peaks. Self-deconvolution, like second derivative spectroscopy, enhances noise, interference fringes, impurity bands and water vapor bands, as well as the true bands of the sample. A high S/N ratio, high nominal instrument resolution, and well-purged instruments are, therefore, essential for good results.

The interpretation of deconvolved spectra is quite analogous to the interpretation of the second derivative spectra discussed in the previous section. For instance, ribonuclease (Fig. 14.1) has an approximate α-helix content of 22% and a β-structure content of about 46% (14). The 1636 cm^{-1} band can be associated with β-segments and the much weaker band near 1655 cm^{-1} with helical segments (5,8). The weak peaks between 1660 and 1700 cm^{-1} evidently correspond to the second β-structure band and to turns. The weak feature at 1646 cm^{-1} may

TABLE 14.1

Assignment of *Amide I* Band Components to Different Types of Protein Conformation

1624 ± 4	Extended chains	1663 ± 4	Turns
1632 ± 2	Extended chains	1669 ± 2	Turns
1638 ± 2	Extended chains	1675 ± 4	Extended chains
1644 ± 3	Unordered	1683 ± 2	Turns
1654 ± 3	Helix	1688 ± 2	Turns
		1694 ± 2	Turns

TABLE 14.2a

Quantitative Secondary Structure Analyses of α-Chymotrypsin

Frequency (cm^{-1})	Assignment	Percent of Total *Amide I* area
1688	Turn	2.3
1680	Turn	5.7
1673	Extended chain	5.4
1665	Turn	15.5
1655	Helix	11.7
1647	Unordered	15.3
1638	Extended chain	25.7
1630	Extended chain	18.4

represent disordered structure.

In the spectrum of α-chymotrypsin, the bands close to 1628, 1637 and 1674 cm^{-1} can be associated with β-strands, the band near 1653 cm^{-1} with the α-helix, and the bands close to 1687, 1681 and 1665 cm^{-1} with turns (9,12). Although a precise assignment of the many amide I components observed after deconvolution (Fig. 14.1) or by second derivative spectroscopy (Figs. 14.2 and 14.5) appears at first sight bewildering, a quite well substantiated general scheme can be obtained on the basis of previous work with different proteins (1–3,5,8,9,12), and a comprehensive study of the deconvolved spectra and second derivative spectra of over twenty proteins recently carried out in this laboratory. Table 14.1 presents these assignments of *amide I* components. Each protein does not exhibit components at each frequency. In particular, extended chains have from one to three components in the 1620-1640 cm^{-1} region and one weak band between 1670 and 1680 cm^{-1}. Helical segments result in a single observable component at 1654 ±3 cm^{-1}. Turns usually exhibit bands close to 1663 cm^{-1} and, in addition, may also have bands around 1669, 1683, 1688 and 1694 cm^{-1}. Unordered, hydrated segments (2,3) produce a characteristic band near 1644 cm^{-1}.

To obtain quantitative structural information, the areas of the components must be measured. This is most easily accomplished with the help of some iterative technique. In this laboratory, Gauss-Newton iteration assuming Gaussian band envelopes has been successfully employed (9,12). The procedure is illustrated by the example of α-chymotrypsin (Fig. 14.1). The pertinent data are given in Table 14.2. If all the percentages of components associated with a particular conformation are added, the total percentage is obtained. For instance, the relative areas of all the components associated with extended chains add up to 49.5% of the total *amide I* area. The best β-structure value from X-ray data (14) is 50%. By contrast, circular dichroism (CD) data vary from 29-50% (15,16).

Table 14.2b gives the percentages of helix, extended chains, unordered segments, and turns as calculated from the data in Table 14.2a for α-chymotrypsin. Literature data obtained from X-ray crystallography (14) and circular dichroism spectroscopy (15,16) are also given for comparison. The later two methods generally provide meaningful results only for helical segments and extended segments.

For "turns" no usable X-ray values are available and only very incomplete CD data exist. The "unordered" fraction, as determined by

TABLE 14.2b

Quantitative Secondary Structure Estimation of α-Chymotrypsin by Different Methods

Conformation	FTIR[a]	X-ray[b]	Circular Dichroism[c]
% Helix	11.7	10	9; 6
% Extended chains	49.5	49	29; 50
% Unordered	15.3	—	—
% Turns	23.5	—	22; 3

[a]Area of all bands assigned to a given conformation as a percentage of the total *Amide I* band area.
[b]Evaluated by Levitt and Greer (14) from original data by Tulinsky et al. (17).
[c]First value from (15), second from (16).

TABLE 14.3

Comparison of Helix Content and Extended Chain Content as Observed by FTIR[a] and X-ray Methods[14]

Protein	Helical Content (%)		Extended chain content (%)	
	FTIR	X-ray	FTIR	X-ray
Carboxypeptidase	40	39	33	30
Chymotrypsinogen	13	11	49	46
Concanavalin A	4	2	60	60
Lysozyme	41	45	21	19
Papain	27	29	32	29

[a]Area of all bands assigned to a given conformation as a percentage of the total *Amide I* band area.

FTIR, refers to the area of the 1644 cm^{-1} region band (see Table 14.1) and most probably represents hydrated chains (2,3).

Table 14.3 gives a comparison between the FTIR values and the X-ray values concerning the helix content and extended chain content for five more proteins studied in this laboratory. We have chosen values as given by Levitt and Greer (14) for X-ray data because their study presents, to the best of our knowledge, the most self-consistent and systematic evaluation of the conformation of a large number of crystallographically studied proteins. The proteins listed in Table 14.3 were selected to present a wide range of helix content and extended chain content. The agreement between X-ray results and FTIR results is generally good. It is, indeed, much better than the agreement between X-ray data–which necessarily must serve as a standard–and other spectroscopic techniques, such as circular dichroism spectroscopy (15,16) or Raman spectroscopy (18,19). (These are based essentially on statistical correlations between broad, poorly resolved spectra and the known structures of a selected set of standard polypeptides or proteins.)

It must be strongly emphasized that the percentage of a particular substructure, such as helix or extended chains, is not a quantity that even in principle can be precisely determined by any method. This is not the fault of the methods, but depends on the nature of protein structure itself. It is not possible, for instance, to say unequivocally where a particular helical segment begins or ends. The agreement for helix

content and extended chain content obtained from X-ray data by Levitt and Greer (14) with the help of sophisticated computer programs, and that obtained in this laboratory by essentially experimental soectroscopic procedures on the other hand, is therefore better than one might expect. No similar comparison is given for "turns" in these proteins because they are not clearly defined by Levitt and Greer (14) (or by anybody else). Nor is a comparison given with circular dichroism or Raman data because no systematic studies by these methods appear to be available. The CD results, furthermore, vary widely from one author to another, as seen in Table 14.2b for α-chymotrypsin.

FTIR spectroscopy thus appears to be as good as, if not better than, any other available technique for estimating the conformation of proteins in aqueous solution. The fact that deuterium oxide must be used as a solvent presents an inconvenience, but is not expected to influence the results concerning conformation (2,3,5,9,12).

Solvent denaturation. FTIR spectra, resolution enhanced by second derivative techniques and Fourier self-deconvolution, also permit quite detailed conformational studies related to solvent denaturation (9,13). The deconvolved spectrum, the original FTIR spectrum, and the second derivative spectrum of native bovine chymotrypsinogen A in D_2O solution, and the corresponding spectra as obtained in 60% (v/v) O-deuterated methanol (MeOD) in D_2O are shown in Figure 14.6. The spectrum of the native protein (Fig. 14.6a) which contains about 45% β-structure and 11% α-helix (14) is somewhat similar to the ribonuclease spectrum shown in Figure 14.1. Extended structure bands are observed at 1636 (strong) and 1674 (weak) cm^{-1}. The 1647 cm^{-1} band is probably associated with unordered sections. The 1654 cm^{-1} helix band appears as a shoulder. The remaining bands are likely due to turns. The denatured protein (Fig. 14.6b) exhibits bands at 1617 cm^{-1} (strong) and 1686 cm^{-1} (weak). Quite similar results are obtained in isopropanol-d solution. The bands at 1617 and 1686 cm^{-1} correspond to no conformation previously studied by infrared spectroscopy. They are, nevertheless, close to usually observed β-structure bands and probably indicate the presence of extended strands of some kind. The ill-defined, broad adsorption between the two sharp bands could be associated with unordered hydrated segments and/or irregular helices (13).

The usefulness of FTIR spectroscopy as a tool for denaturation studies is further illustrated in Figure 14.7, which shows the spectra of the native form and three different denatured forms of the protein β-lactoglobulin A. The spectrum of the native material (Fig. 14.7a), is typical for a protein with a high extended-structure content; the assign-

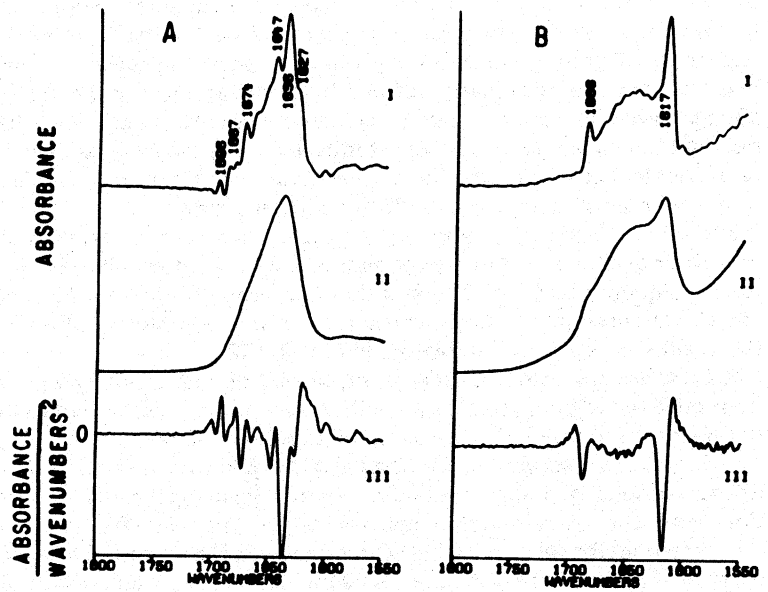

Fig. 14.6. FTIR spectra of chymotrypsinogen A in two different solvents: (I) deconvolved spectrum [deconvolution constants: σ = 6.5 cm^{-1}, K = 2.4]; (II) original spectrum; (III) second derivative of original spectrum. (A) Native protein in D_2O, 4.7% w/v, pD 7; (B) in 60% v/v MeOD solution in D_2O, 0.47% w/v protein at pD 3. [From ref. 13].

ment was discussed in conjunction with the second derivative spectrum shown in Figure 14.5b. The spectrum of alkaline-denatured β-lactoglobulin A at pD 13 is shown in Figure 14.7b. A single strong, broad band is observed around 1640 cm^{-1}, close to the characteristic frequency for unordered segments (Table 14.1), although the unsymmetrical shape suggests that some order is retained and, therefore, the band center is below the typical value of 1644 ± 3 cm^{-1}. The spectrum obtained in acidic MeOD solution is shown in Figure 14.7c. Here the strong band at 1647 cm^{-1} could be associated with helical segments which are reported to prevail under these conditions (20). The band is observed at a lower frequency than the helix band in aqueous solution (1654 ± 3 cm^{-1}; Table 14.1) possibly because of the influence of the different solvent, MeOD. The weak bands at 1618 and 1687 cm^{-1} are associated

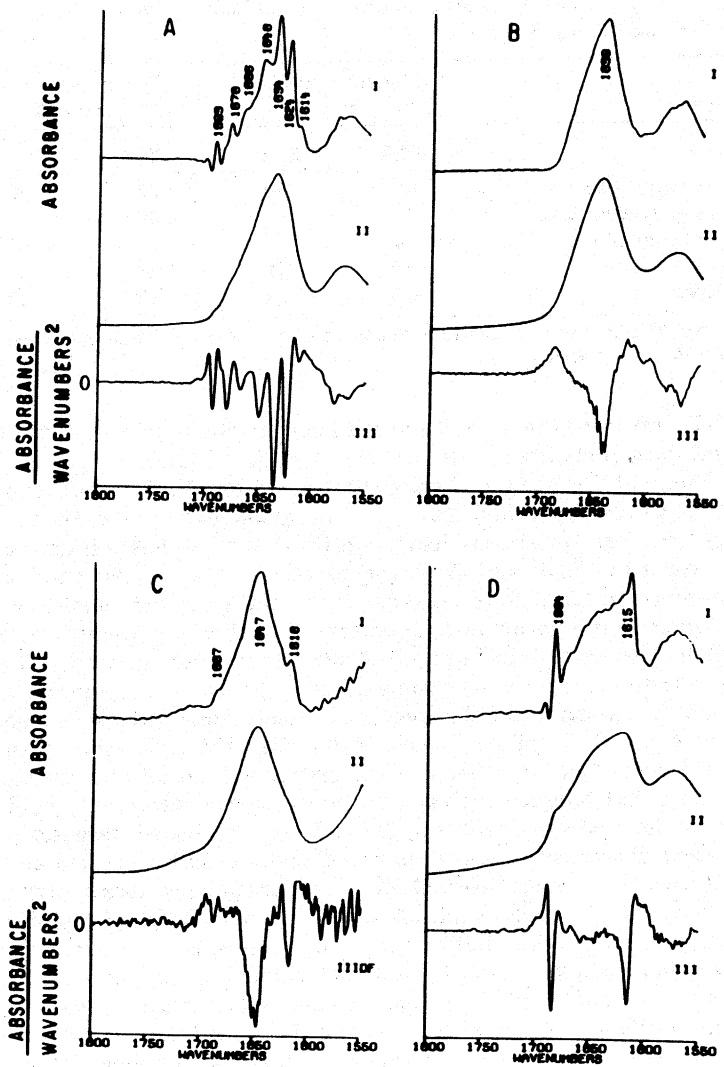

Fig. 14.7. FTIR spectra of β-lactoglobulin A in various solvents: (I), (II), (III) as in Fig. 6. (A) Native protein in D_2O, 5% w/v, pD 7; (B) in D_2O, 5% w/v, denatured at pD 13; d(C) in 60% v/v MeOD in D_2O, 0.4% w/v protein, pD 4. (D) In 40% v/v isopropanol-*d* in D_2O, 0.35% w/v protein, pD 7. [From ref. 13].

with a conformation similar to the one observed for chymotrypsinogen A under similar conditions (Fig. 14.6b). It is interesting to note that acidic MeOD has a different effect on the two proteins and that this difference is clearly reflected in the spectra obtained. The sharp peaks in Figure 14.7c below 1600 cm^{-1} are probably caused by noise. They illustrate how careful one must be with resolution-enhanced spectra.) In isopropanol-*d* solution, as shown in Figure 14.7d, β-lactoglobulin A behaves like chymotrypsinogen A.

Figure 14.7 thus shows how FTIR spectra can clearly distinguish between the native form and three denatured forms of the same protein. We know of no other spectroscopic technique that will accomplish this in quite as much detail.

FTIR spectroscopy permits much more extensive protein structure studies than the older, dispersive infrared techniques. The most comprehensive results are obtained by applying Fourier self-deconvolution and second derivative spectroscopy. Both qualitative and quantitative conformation studies are possible by these techniques. In aqueous solution, more detailed information is obtained than by any other spectroscopic method, although D_2O must be used as a solvent in order to obtain good results. It is also possible to investigate changes in protein conformation caused by environmental and other causes, such as different solvents and variations in pH. The results are more specific than the ones obtained by other spectroscopic methods.

References

1. Elliott, A., and E.J. Ambrose, *Nature 165*:921 (1950).
2. Timasheff, S.M., H. Susi and L. Stevens, *J. Biol. Chem 242*:5467 (1967).
3. Rüegg, M., V. Metzger and H. Susi, *Biopolymers 14*:1465 (1975).
4. Koenig, J.L., *Adv. Polymer Sci. 54*:87 (1984).
5. Yang, W.-J., P.R. Griffiths, D.M. Byler and H. Susi, *Appl. Spectrosc. 39*:282 (1985).
6. Kauppinen, J.K., D.J. Moffatt, H.H. Mantsch and D.G. Cameron, *Ibid. 35*:271 (1981).
7. Kauppinen, J.K., D.J. Moffatt, H.H. Mantsch and D.G.Cameron, *Anal. Chem. 53*:1454 (1981).
8. Susi, H., and D.M. Byler, *Biochem. Biophys. Res. Comm. 115*:391 (1983).
9. Susi, H., and D.M. Byler, in "Methods in Enzymology," edited by C.H.W. Hirs and S.N. Timasheff, Academic Press, Inc., New York, 130:290 (1986).
10. Yang, W.-J., and P.R. Griffiths, *Proc. Soc. Photo-Opt. Instrum. Eng. 289*:263 (1981).

11. Koenig, J.L., and D.L. Tabb, in "Analytical Applications of FT-IR to Molecular and Biological Systems," edited by J.R. Durig, D. Reidel Publishing Co., Dordrecht, Holland, 1980, p. 241.
12. Byler, D.M., and H. Susi, *Biopolymers 25*:469 (1986).
13. Purcell, J.M., and H. Susi, *J. Biochem. Biophys. Meth. 9*:193 (1984).
14. Levitt, M., and J. Greer, *J. Mol. Biol. 114*:181 (1977).
15. Provencher, S.W., and J. Glockner, *Biochemistry 20*:33 (1981).
16. Chang, C.T., C.-S. Wu, Y.T. Yen, *Anal. Biochem. 91*:13 (1978).
17. Tulinshky, A., N.V. Mari, C.N. Morimoto and R.L. Vandler, *Acta Crystallogr; B29*:1309 (1973).
18. Berjot, M., J. Marx, and A.J.P. Alix, *J. Raman Spectrosc. 18*:289 (1987).
19. Williams, R.M., *Meth. Enzymol. 130*:311 (1986).
20. Inoue, H., and S.N. Timasheff, *J. Am. Chem. Soc. 90*:1890 (1968).